

Dataset Reviewer's Guide

June 2023

Copyright, 2023 Behavox - This document and its content are protected by Canadian copyright law. Except as otherwise provided for under Canadian copyright law, this document and its content may not be copied, published, distributed, downloaded or otherwise stored in a retrieval system, transmitted or converted, in any form or by any means, electronic or otherwise, without the prior written permission of the copyright owner.

Contents

- Dataset prep.....1**
 - Datalake registration.....3
 - Registering on the datalake.....3
 - Revoking a datalake registration.....4
 - Dataset configuration files.....5
 - Uploading a dataset configuration file to your Behavox Compliance instance.....5
 - Dataset reviews.....6
 - Reviewing a dataset.....6
 - Bookmarking a dataset.....8
 - Downloading and sending a dataset by email.....8
 - Uploading a dataset.....9

Dataset prep

As part of a comprehensive AI feedback program, Behavox Compliance provides a convenient tool set to manage manual dataset review and approval.

Why is dataset prep important ?

Feedback is a powerful way to fine-tune alerts and reduce the workload associated with false positives. The dataset review and approval workflow ensures you have complete control of the content used for ML feedback. The Behavox feedback program is guided by the following principles:

- Centralized – All training and retraining is performed centrally by Behavox with no local training of models on client environments.

Benefits:

- Delivers consistency in retraining – avoids sharding.
- Careful curation of data for retraining.
- Centralization benefits all clients by improving diversity of training data.
- Standardized Risk documentation.
- Controlled – Retraining runs through a formal release process, and there is no automatic retraining of models in production based on client escalations or alert closures.

Benefits:

- Ensures that no updates to the production models occur without careful curation of training data and evidence of the performance improvement from the change.
- Retrained models must meet acceptance criteria (same or improved F1 score) to release.
- Continuous – Behavox ML feedback program is a long-term investment in model improvement. Data collection occurs at least quarterly for all models.

Dataset prep configuration

Behavox sends you a dataset configuration file that targets specific attributes like:


- ME groups
- Date range
- Data types
- Target number of sentences

Dataset prep workflow

The dataset review workflow consists of the following steps:

1. Behavox creates a JSON task configuration file and sends it to you.
2. You upload and run the file in Behavox Compliance.
3. Behavox prepares a dataset. Samples are created and the content is anonymized.
4. You review the content. Remove or redact dataset text at your discretion.
5. You approve the dataset.
6. You upload the dataset directly to the data lake, or you download the dataset and email it to Behavox.



 **Note:** Uploading the dataset requires a Behavox admin to configure access to the data lake.

The following entities are automatically anonymized in samples:

- Organization
- Person name
- Email
- Phone number
- Signature tag

Effective collaboration on dataset prep

Bookmarking is a great way for teams to work on a dataset together. You can place a bookmark in a dataset to mark the extent of your progress in a review. Bookmarks ensure no duplication of effort when you or teammates resume the review.

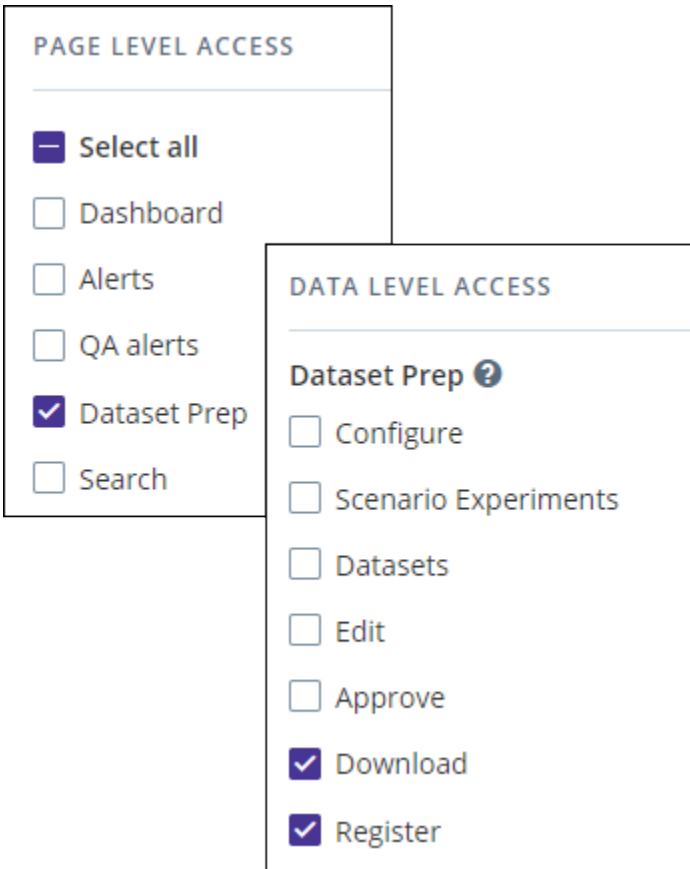
Datalake registration

The dataset prep workflow supports an option to upload an approved dataset directly from Behavox Compliance, instead of emailing the dataset to Behavox.

This enhancement makes providing ML feedback much more convenient and secure.

 **Note:** Datalake registration requires an access key. Contact Behavox for a key.

You need the **Dataset Prep>Register** permission for your role to enter this key in Dataset Prep. Once the Datalake registration is successful, the **Dataset Prep>Download** permission enables the upload functionality for users.



The image shows a user interface for configuring permissions. It consists of two overlapping panels. The left panel, titled 'PAGE LEVEL ACCESS', has a 'Select all' button and a list of items: 'Dashboard', 'Alerts', 'QA alerts', 'Dataset Prep' (which is checked), and 'Search'. The right panel, titled 'DATA LEVEL ACCESS', is open and shows a list of actions under the 'Dataset Prep' header: 'Configure', 'Scenario Experiments', 'Datasets', 'Edit', 'Approve', 'Download' (checked), and 'Register' (checked).

PAGE LEVEL ACCESS	
<input checked="" type="checkbox"/>	Select all
<input type="checkbox"/>	Dashboard
<input type="checkbox"/>	Alerts
<input type="checkbox"/>	QA alerts
<input checked="" type="checkbox"/>	Dataset Prep
<input type="checkbox"/>	Search

DATA LEVEL ACCESS	
Dataset Prep ?	
<input type="checkbox"/>	Configure
<input type="checkbox"/>	Scenario Experiments
<input type="checkbox"/>	Datasets
<input type="checkbox"/>	Edit
<input type="checkbox"/>	Approve
<input checked="" type="checkbox"/>	Download
<input checked="" type="checkbox"/>	Register

Once you apply your access key, all dataset reviewers will be able to upload datasets. Keys are not specific to a profile, but are subject to permissions you have configured for your roles.

Registering on the datalake

Datalake registration enables all dataset reviewers to upload approved datasets to Behavox, instead of emailing them.

Before you begin

 **Note:** Datalake registration requires an access key. Contact Behavox for a key.

About this task

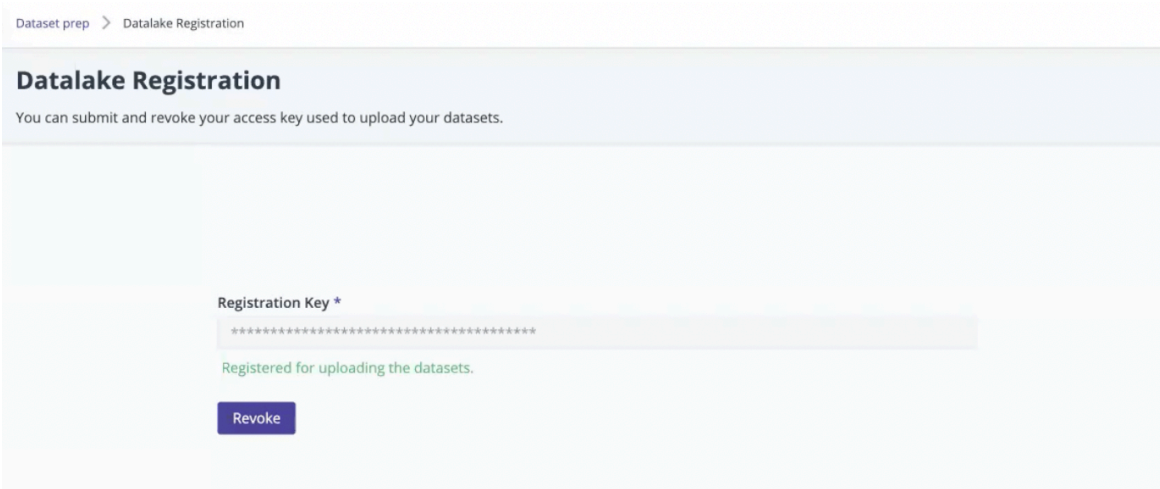
At your Behavox Compliance instance:

Procedure

1. Open the **Datalake registration** page. Click **Dataset Prep > Datalake registration**.
2. Paste your Behavox-provided access key into the **Registration Key** dialog box.
3. Commit your changes. Click **Save**.

Results

The system displays a success message.




The screenshot shows the 'Datalake Registration' page. At the top, there is a breadcrumb 'Dataset prep > Datalake Registration'. Below this is a header 'Datalake Registration' with a subtext 'You can submit and revoke your access key used to upload your datasets.' The main content area has a light blue background. In the center, there is a form with a label 'Registration Key *' and a text input field containing a series of asterisks. Below the input field, a green message states 'Registered for uploading the datasets.' At the bottom of the form, there is a purple button labeled 'Revoke'.

Revoking a datalake registration

You can revoke a datalake registration and force all dataset reviewers to email approved datasets to Behavox, instead of uploading them.

About this task

 **Note:** Once you revoke the access key, all dataset reviewers will be unable to upload datasets.

At your Behavox Compliance instance:

Procedure

1. Open the **Datalake registration** page. Click **Dataset Prep > Datalake registration**.
2. Revoke the registration. Click **Revoke**.

Dataset configuration files

Dataset configuration files are a convenient way to automate the collection and packaging of targeted data.

Behavox sends you a dataset configuration file that targets specific attributes like:

- ME groups
- Date range
- Data types
- Target number of sentences

This file is sent to you through email.

To extract sample text, you need to upload the dataset configuration file to your Behavox Compliance instance.

Uploading a dataset configuration file to your Behavox Compliance instance

Upload your new dataset configuration file to your Behavox Compliance instance to create sample text.


About this task

At your Behavox Compliance instance:

Procedure


1. Open the **Datasets Upload** page. Click **Dataset Prep > Dataset Upload**.
2. Upload and run your dataset configuration file.
 - a. Open the upload tool. Click the **Upload** button.
 - b. Select the dataset configuration file sent to you by Behavox. Click **Choose File** and browse.
 - c. Submit your file to Behavox Compliance. Click **Submit**.
The file runs automatically.

When you dataset is uploaded, the status column will display a green check mark.

 **Note:** If an error occurs that prevents the dataset from uploading, the status column will display an error message link. Click the link to view details for the error. Usually, errors must be corrected by Behavox. Contact your Customer Success Manager and relay the error details.


Dataset reviews

You must review and approve the sample text before the dataset can be released to Behavox.

 **Note:** Datasets are automatically saved while you review. You can resume a review anytime without losing your work.

The **Datasets Review** page contains a listing of reviews that are either **In review** or **Approved**. You can check the status of a review in the **Status** column of the listing:

- **Preparing** – Datasets are disabled while the extraction task is running. This is a transient state and the dataset will be **In review** shortly.
- **In review** – Datasets require approval.
- **Approved** – Datasets are reviewed.
- **Uploaded** – Datasets have been sent to the data lake.

 **Note:** The datasets page will have additional columns with uploads details, if your system is configured with a data lake integration.

Reviewing a dataset

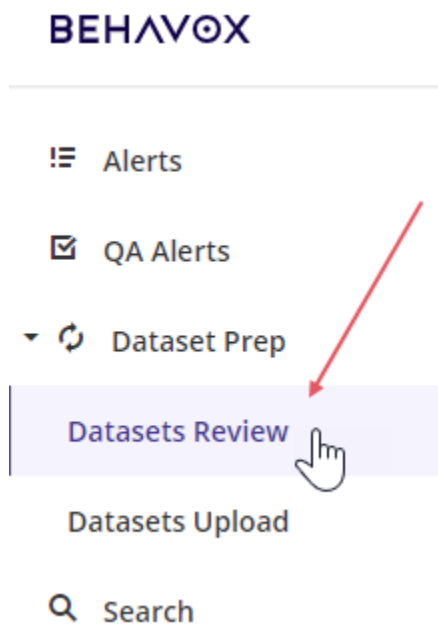
You can open a dataset for review from the **Datasets** page.

About this task

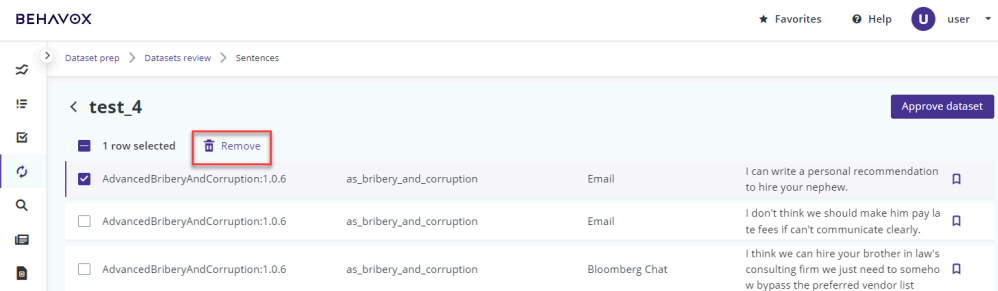
At your Behavox Compliance instance:

Procedure

1. Open the **Datasets** page. Click **Dataset Prep > Datasets Review**.



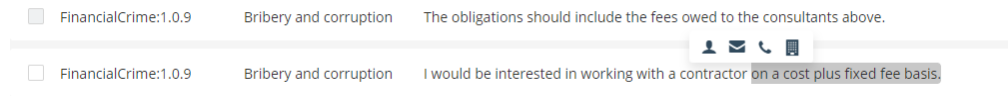
2. Open the Dataset. Click the dataset from the list.
3. Optional: Remove an entire entry from the dataset.
 - a. Select the entry to remove. Check the check box for the entry.
 - b. Remove the entry. Click the trash bin.



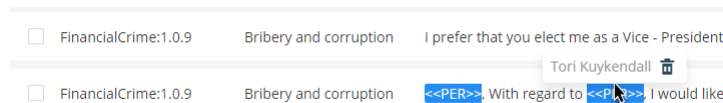
Note: You can make multiple selections when you delete.

4. Optional: Redact text from an entry.

- a. Select the text to redact. Click and drag to highlight the text.
- b. Select a redaction attribute from the redaction tool. Choose from: **Person**, **Email address**, **Phone number**, or **Organization**.



You can reveal redacted text by clicking the redaction. You can suspend a redaction by clicking it and selecting the trash can.



5. If you have no further changes, approve the dataset. Click **Approve**.

📌 **Important:** You cannot change an approved dataset. Ensure your review is complete before you proceed.

Bookmarking a dataset


You can place a bookmark in a dataset that you are reviewing, so that you can resume your review at a later time without losing your place.

About this task

At your **Dataset** page:

Procedure

Mark an entry in the dataset. Click the bookmark icon  for the dataset entry.

💡 **Note:** You can update or remove a bookmark by clicking the bookmark icon  for the desired dataset entry.

Downloading and sending a dataset by email


About this task

Download and send your dataset to Behavox to provide AI feedback.

💡 **Note:** You must approve a dataset before you can download it from Behavox Compliance. Complete [Reviewing a dataset](#) if your dataset is not approved.

At your Behavox Compliance instance:


Procedure


1. Open the **Datasets** page. Click **Dataset Prep > Datasets Review**.
2. Select the dataset you want to send to Behavox. Click the dataset item.
3. Download the dataset. Click the **Download**  button.
4. Send the dataset file to Behavox with your mail client, using the address provided to you.

Uploading a dataset

Upload your dataset to Behavox to provide AI feedback.

Before you begin

 **Note:** You must complete [Registering on the datalake](#) before you can upload a dataset.

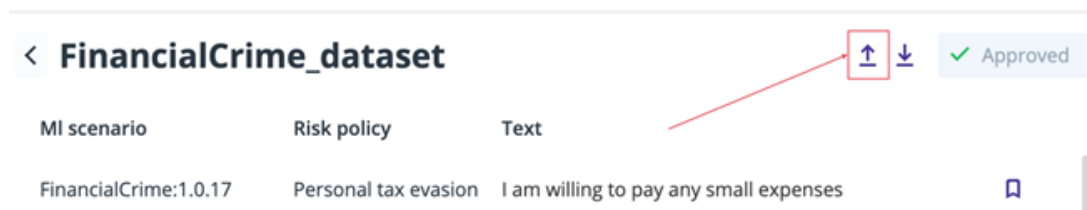
 **Note:** You must approve a dataset before you can upload it to Behavox Compliance. Complete [Reviewing a dataset](#) if your dataset is not approved.

About this task

At your Behavox Compliance instance:

Procedure

1. Open the **Datasets** page. Click **Dataset Prep > Datasets Review**.
2. Select the dataset you want to send to Behavox. Click the dataset item.
3. Upload the dataset. Click the **Upload** button.



www.behavox.com

BEHAVOX